Audio Quality Research

Outputs

- Technical publications and presentations on new research results.
- Measurements and estimates of speech and audio quality and algorithm performance.
- Algorithms and data supporting speech and audio coding and quality assessment.

Digital coding and transmission of speech and audio signals are enabling technologies for many telecommunications and broadcasting services including cellular telephone services, voice over Internet protocol (VoIP) services, and digital audio broadcasting systems. Speech and audio signals can be coded and transmitted at low bit-rates with good fidelity. In addition, coded speech and audio signals can be packetized for transmission, thus sharing radio spectrum or wired network bandwidth with other data streams and hence with other users.

Innovation in digital coding and transmission involves compromises and trade-offs among speech or audio quality, transmission bit-rate, robustness to transmission errors and losses, coding and transmission delay, and coding and transmission algorithm complexity. The ITS Audio Quality Research Program works to identify and develop new techniques to increase quality or robustness, or to lower bit-rate, delay, or complexity of digital speech and audio coding and transmission algorithms. The ultimate result of such advances is better sounding, more reliable, more efficient communications and broadcasting services.

The robustness of digital coding and transmission algorithms is critical in applications that use lossy channels such as those associated with wireless systems and those provided by the Internet. In FY 2005, Program staff have continued to work towards more robust speech coding through a method called multi-descriptive coding (MDC). In MDC an encoder forms multiple partial descriptions of a speech signal and these descriptions are sent over different channels. If all descriptions arrive at the decoder intact, a higher-quality reconstruction of the speech is possible. If channel failures cause any of the descriptions to be lost, then a lower-quality reconstruction of the speech signal is still possible.

Program staff developed and tested a multiple description pulse code modulation (PCM) speech coding system that exploits naturally occurring correlations between adjacent samples of speech and invokes a pair of appropriately designed vector quantizers. This system includes an aspect ratio parameter that allows one to trade off the speech quality when one channel is working against the speech quality when two channels are working. This in turn allows one to match the channel conditions to maximize speech quality.

Even when robust coding techniques such as MDC are deployed, time-varying channel conditions generally will cause end users to experience time-varying speech quality. But the perception of time-varying speech quality is not yet well-understood. For example, which would be preferable: speech that fluctuates between high quality and low quality, or speech that is consistently of medium quality? How would the preference change with the levels of high, medium, and low quality? And how would the preference change with the timing and nature of the quality fluctuations?

Program staff recently designed, conducted, and analyzed an experiment to characterize one simple yet fundamental component of time-varying speech quality. In this experiment, subjects heard recordings where speech quality changed twice, resulting in speech quality histories of the form "low, high, low" and "high, low, high." Figure 1 below shows an example result from this experiment. The vertical

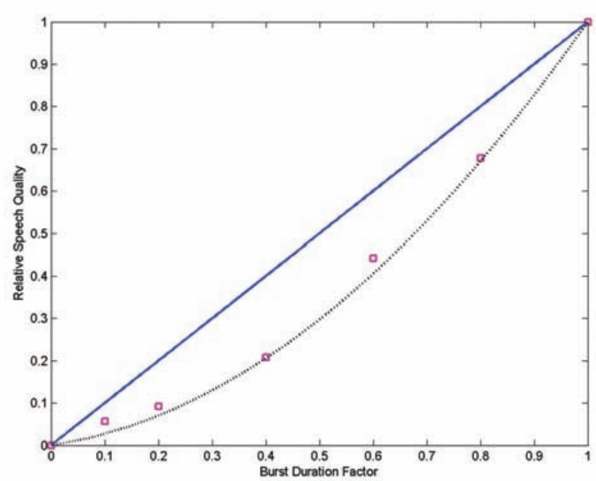

Figure 1. Example results from an experiment on time-varying speech quality. See text for details.

axis represents overall speech quality as judged by the subjects in the experiment. Here 0 represents low speech quality and 1 represents high speech quality. On the horizontal axis, the burst duration factor describes the fraction of the time that the recording has speech quality 1. (The remainder of the time the recording has speech quality 0.) If subjects judged overall speech quality by simply averaging instantaneous speech quality, then their responses would grow linearly as described by the solid blue line in the figure. Instead, subjects' responses fall below this line, as denoted by purple squares and approximated by the dotted black line in the figure. This indicates that subjects use a process that is more critical than simple averaging when judging overall speech quality. Periods of low and high quality do not balance out in the mathematical sense; rather, periods of low quality seem to carry more weight than periods of high quality. In this sense, we might say that listeners are pessimists when it comes to speech quality.

Figure 2. Subjective speech and audio testing can now be controlled by test participants using wireless PDAs (photograph by S. Wolf).

In FY 2005, Program staff designed, installed, and tested upgrades and new capabilities to the ITS Audio-Visual Laboratories. One major upgrade provides greatly enhanced flexibility for subjective test design and operation, in that subjective tests can now be controlled through a very flexible, powerful, and easy to use, high level language that is well-integrated with digital and analog audio I/O, computer graphics display, text display, text entry, and mouse input. Graphical display and mouse input is extended from a host desktop computer (operated by Program staff) to a handheld personal digital assistant (PDA) (operated by a subjective test participant). This extension is enabled by a wireless local area network (LAN) so that test participants only need to operate a single familiar, intuitive, lightweight, wireless device (see Figure 2 above).

Throughout FY 2005, program staff continued with subjective and objective audio quality testing to support this and other ITS programs. Staff continued to transfer technologies to industry, Government, and academia through numerous technical publications, presentations, guest lectures, laboratory demonstrations, and by completing peer reviews for technical journals and workshops. Program staff also incorporated recent program results into a revision of an American National Standard: ANSI T1.801-04. This telecommunications standard is titled "Multimedia Communications Delay, Synchronization, and Frame Rate," and it now includes the specification of an ITS-developed algorithm for tracking variable transmission delay across a wide range of speech coding conditions. Program publications, technical information, and other program results are available at http://www.its.bldrdoc.gov/audio.

Recent Publications

S.D. Voran, "A multiple-description PCM speech coder using structured dual vector quantizers," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, Philadelphia, Mar. 2005.

S.D. Voran, "Multiple-description PCM speech coding by complementary asymmetric vector quantizers," in *Proc. IEEE Region 5 Conference*, Boulder, CO, Apr. 2005.

S.D. Voran, "A basic experiment on time-varying speech quality," in *Proc. 4th International MESAQIN (Measurement of Speech and Audio Quality in Networks) Conference*, Prague, Czech Republic, Jun. 2005.

For more information, contact:
Stephen D. Voran
(303) 497-3839
e-mail svoran@its.bldrdoc.gov